Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2013)
62, *Part* 3, *pp.* 451–472

Reduced hierarchical models with application to estimating health effects of simultaneous exposure to multiple pollutants

Jennifer F. Bobb and Francesca Dominici

Harvard School of Public Health, Boston, USA

and Roger D. Peng

Johns Hopkins School of Public Health, Baltimore, USA

[Received January 2012. Revised October 2012]

Summary. Hierarchical models (HMs) have been used extensively in multisite time series studies of air pollution and health to estimate health effects of a single pollutant adjusted for other pollutants and other time varying factors. Recently, the US Environmental Protection Agency has called for research quantifying health effects of simultaneous exposure to many air pollutants. However, straightforward application of HMs in this context is challenged by the need to specify a random-effect distribution on a high dimensional vector of nuisance parameters. Here we introduce the *reduced HM* as a general statistical approach for analysing correlated data with many nuisance parameters. For reduced HMs we first calculate the integrated likelihood of the parameter of interest (e.g. the excess number of deaths attributed to simultaneous exposure to high levels of many pollutants), and we then specify a flexible random-effect distribution directly on this parameter. Simulation studies show that the reduced HM performs comparably with the full HM in many scenarios and even performs better in some cases, particularly when the multivariate random-effect distribution of the full HM is misspecified. Methods are applied to estimate relative risks of cardio-vascular hospital admissions associated with simultaneous exposure to elevated levels of particulate matter and ozone in 51 US counties during 1999–2005.

Keywords: Air pollution; Multilevel models; Multisite time series data; Nuisance parameters; Random effects

1. Introduction

The US Environmental Protection Agency estimated that thousands of premature deaths and hundreds of thousands of cases of illness may be avoided by reducing pollution (Environmental Protection Agency, 2011). Most epidemiological studies of air pollution and health have estimated the health effects that are associated with ambient exposure to individual pollutants adjusted for exposure to other pollutants and confounders. However, the National Research Council has recently questioned whether the current approach of setting separate national ambient air quality standards (NAAQSs) for each of the six criteria pollutants adequately protects population health, as this approach may greatly underestimate risk (National Research Council, 2004). To meet the challenges of the National Research Council recommendations, new statistical methods are needed to account for multiple exposures and their interactions.

Address for correspondence: Jennifer F. Bobb, Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA.
E-mail: jbobb@hsph.harvard.edu

 0035–9254/13/62451

Previous multisite time series studies of the health effects of air pollution have estimated risks that are associated with exposure to a single pollutant. Dominici *et al.* (2000) developed a two-stage hierarchical model to combine information across locations on the association between daily changes of a given pollutant and daily changes in the health outcome, adjusted for other pollutants and confounders. This approach has been applied to several national US studies for estimating independent associations of various pollutants of epidemiologic interest with different health outcomes, including mortality and cardio-vascular and respiratory emergency hospital admissions (Dominici *et al.*, 2006; Bell *et al.*, 2004; Peng *et al.*, 2008, 2009). Two-level random-effect models have also been used to estimate health effects of exposure to individual pollutants and to identify factors that explain heterogeneity in the health risks across European cities (Katsouyanni *et al.*, 2001). Addressing the potential for biased estimates due to measurement error of correlated exposures in multipollutant models, Zeka and Schwartz (2004) applied methodology that was developed by Schwartz and Coull (2003) to estimate independent effects of individual pollutants that minimizes the effect of measurement error.

To estimate the health effects of simultaneous exposure to multiple pollutants, we specify a hierarchical model (HM) that, at the first stage, flexibly specifies the air pollution–health outcome risk surface by incorporating interactions between pollutants and allowing for smooth non-linear functions of pollutant concentrations. In the full HM, we define $\boldsymbol{\beta}_i$ to be the random effects describing the association between the health outcome and the multiple exposure variables that are included in the regression model (e.g. non-linear functions of main effects and interactions of pollution variables and potential confounders) for the ith location. The parameter of primary scientific interest (θ_i) is the increased health risk when daily ambient levels of the pollutants considered are simultaneously above their national standards compared with when daily levels are below their national standards. Our goals are to obtain more precise estimates of θ_i by borrowing strength across locations, to estimate overall regional or national risks θ^* and to identify site-specific factors (e.g. population demographics, traffic patterns and long-term averages of other pollutants) that modify the association between simultaneous exposure to multiple pollutants and adverse health outcomes.

More generally, the hierarchical modelling approaches that we consider apply to problems where the parameter of interest θ_i can be defined as a known function of $\boldsymbol{\beta}_i$ where $\dim(\boldsymbol{\beta}_i) \gg \dim(\theta_i)$. Many difficulties may arise on implementation of standard generalized linear mixed models or full HMs in the presence of a high dimensional vector of random effects ($\boldsymbol{\beta}_i$). First, one must specify a multivariate random-effect distribution on the full vector $\boldsymbol{\beta}_i$, which might not be of primary scientific interest. There is an extensive literature on the consequences of misspecification of random-effect distributions in generalized linear mixed models (Verbeke and Lesaffre, 1997; Heagerty and Kurland, 2001; Litière *et al.*, 2008; Agresti *et al.*, 2004). Though small to moderate misspecification of the random-effect distribution may not have a large effect in the estimation of fixed effects, there are situations for which misspecification can result in a loss of efficiency and biased estimates of the random effects (Neuhaus *et al.*, 1992; Heagerty and Kurland, 2001; Agresti *et al.*, 2004; Litière *et al.*, 2010; McCulloch and Neuhaus, 2011). Several approaches have been proposed for specifying flexible semiparametric or non-parametric distributions for the random effects (Laird, 1978; Magder and Zeger, 1996; Komárek and Lesaffre, 2008; Gallant and Nychka, 1987; Chen *et al.*, 2002). However, most of these approaches cannot be implemented in the context of a high dimensional vector of random effects, and the validity of the assumption on the random-effect distribution is sometimes difficult to verify (Agresti *et al.*, 2004; Litière *et al.*, 2008). Second, if one is interested in estimating effect modification, at the second stage the

full HM presents the additional challenge of specifying a high dimensional multivariate regression model. Third, implementing diagnostic methods for misspecification of a multivariate random-effect distribution can be very challenging. Fourth, it may be computationally intensive and/or challenging to implement a Markov chain Monte Carlo (MCMC) sampler that mixes well and converges quickly to the stationary distribution as the number of random effects increases.

In this paper, we introduce *reduced HMs* as a general statistical approach for eliminating nuisance parameters in HMs with a large number of random effects. The reduced HM combines information across clusters (e.g. locations) directly on the parameter of interest θ_i. At the first stage, we calculate an integrated likelihood for θ_i and, at the second stage, we specify a flexible random-effect distribution directly on the θ_i. Reduced HMs overcome many of the practical challenges in the specification and implementation of full HMs in the context of a high dimensional vector of nuisance parameters. Though developed to study health effects of simultaneous exposure to multiple pollutants, reduced HMs are widely applicable for other studies of multiple exposures, and in general to clustered data sets with a large number of nuisance parameters. Accordingly, much of the methods section is presented in a general context while maintaining a close connection to the scientific motivation for this work.

Previous studies have used likelihoods of the parameter of interest at the first stage of a hierarchical model for conducting a meta-analysis of randomized trials of a treatment for stomach ulcers (Efron, 1996; Liao, 1999). Specifically, Efron (1996) used a conditional likelihood for the clinical-trial-specific log-odds ratio (θ_i) and developed empirical Bayes methods for combining the likelihoods to conduct inference (interval estimation) on the θ_i. Liao (1999) also eliminated nuisance parameters at the first stage by using conditional likelihoods, but he modelled the θ_i by using a Bayesian approach, assuming a normal random-effect distribution for the θ_i. In these two studies, the vector of cluster-specific parameters $\boldsymbol{\beta}_i$ is just two dimensional, and a conditional likelihood for θ_i is available in closed form. Although not explicitly defining a likelihood function to eliminate nuisance parameters at the first stage of the HM, Warn *et al.* (2002), building on the work by Smith *et al.* (1995), reparameterized the cluster-specific parameters $\boldsymbol{\beta}_i$ as $(\boldsymbol{\lambda}_i, \theta_i)$, where θ_i is the parameter of interest, and then proposed to use non-informative priors for the nuisance parameter $\boldsymbol{\lambda}_i$, which were assumed to be independent across clusters. However, it may not always be possible to define such a reparameterization (e.g. if θ_i is a complex function of $\boldsymbol{\beta}_i$), and this approach still requires sampling the nuisance parameter $\boldsymbol{\lambda}_i$ at each iteration of the MCMC algorithm, which can become computationally expensive when the dimension of $\boldsymbol{\lambda}_i$ is large. In this paper we generalize parameter reduction for HMs to very general situations where

(a) no conditional or marginal likelihood is available,
(b) an integrated likelihood is not available in closed form,
(c) a reparameterization $(\boldsymbol{\lambda}_i, \theta_i)$ of the within-cluster parameter space does not exist,
(d) the second-level model includes cluster-specific covariates and
(e) flexible specifications of the random-effect distribution are desired.

This generalization is referred to as a reduced HM. Additionally, although there are several practical advantages of the reduced HM arising from the elimination of nuisance parameters at the first stage, even in the specific context where this approach has been applied previously (two-dimensional settings with conditional likelihood available in closed form), there is a lack of evidence supporting the reduced HM as performing competitively with the full HM across a range of scenarios. To address this gap, we shall provide a critical evalu-

ation of the reduced HM as an alternative to fitting the full HM in a series of simulation studies.

In Section 2, we describe the multisite time series data that are used to estimate the health risks associated with simultaneous exposure to multiple pollutants. In Section 3, we describe the level 1 model of an HM aimed at estimating the association between joint exposure to ozone (O_3) and fine particulate matter and hospital admissions. In Section 4, we introduce the reduced HM in a general setting where an integrated likelihood is estimated for each cluster and a flexible random-effect distribution is specified directly on the cluster-specific parameter of interest. Section 5 describes our simulation study. In Section 6, we present our results from the data analysis. We provide discussion and concluding remarks in Section 7.

2. Data

We used data from a national database consisting of parallel time series from 60 counties in the north-eastern USA during the period 1999–2005. Daily counts of emergency hospital admissions for cardio-vascular diseases (CVDs), which comprise heart failure (international classification of diseases ICD-9 code 428), heart rhythm disturbances (codes 426–427), cerebro-vascular events (codes 430–438), ischaemic heart disease (codes 410–414 and 429) and peripheral vascular disease (codes 440–448) were obtained from billing claims of US Medicare enrollees. CVD admissions were stratified by two age categories: 65–74 and 75 years or older. Concentrations of fine particulate matter, $PM_{2.5}$ (micrograms per cubic metre), and O_3 (parts per billion), which for many counties are measured on either a 1-in-3 or 1-in-6 day schedule, were obtained from the US Environmental Protection Agency's air quality system. Daily temperature and dewpoint temperature were obtained from the National Climatic Data Center. Among the 60 north-eastern US counties with available data, we considered the 51 counties having at least 100 days where $PM_{2.5}$- and O_3 levels were measured concurrently, as well as at least one day when both pollutants were above their national standard (which is defined below). Fig. 1 shows a map of the locations, as well as example time series of $PM_{2.5}$- and O_3 levels for Washington DC.

3. Poisson regression model for multiple pollutants

In this section we describe the first level of an HM for estimating health effects that are associated with simultaneous exposure to fine particulate matter ($PM_{2.5}$) and O_3. We assume for county i on day j for age group k that the number of CVD admissions y_{ijk} has a Poisson distribution with mean model

$$\begin{aligned}\log(\mathbb{E}[y_{ijk}]) = \log(n_{ijk}) + \gamma_{i0} + \mathrm{ns}(\mathrm{PM}_{2.5_{ij}}; 3\,\mathrm{DF}, \mathbf{b}_{i1})\,\mathrm{ns}(\mathrm{O}_{3_{ij}}; 3\,\mathrm{DF}, \mathbf{b}_{i2}) \\ + \gamma_{i1}\,\mathrm{age}_k + \boldsymbol{\gamma}'_{i2}\,\mathrm{dow}_{ij} + \mathrm{ns}(\mathrm{temp}_{ij}; 6\,\mathrm{DF}, \gamma_{i3}) + \mathrm{ns}(\mathrm{dptp}_{ij}; 3\,\mathrm{DF}, \gamma_{i4}) \\ + \mathrm{ns}(\overline{\mathrm{temp}}^{(3)}_{ij}; 6\,\mathrm{DF}, \gamma_{i5}) + \mathrm{ns}(\overline{\mathrm{dptp}}^{(3)}_{ij}; 3\,\mathrm{DF}, \gamma_{i6}) + \mathrm{ns}(j; 7\,\mathrm{DF/year}, \gamma_{i7}),\end{aligned} \quad (1)$$

where n_{ijk} is the number of individuals of the kth age group at risk, and ns$(\cdot)$ denotes natural cubic splines with the specified degrees of freedom (DFs) and $\mathbf{b}_{ij}$ $(j = 1, 2)$ and γ_{ij} $(j = 3, \ldots, 7)$ representing the spline coefficients. The product of the cubic spline bases for $PM_{2.5}$ and O_3, which includes both main effects and interaction terms, provides a flexible specification of the unknown joint pollutant–hospital admissions exposure–response surface. Here age denotes an indicator for being in the 75 years and older age category (*versus* 65–74 years), dow is a vector of indicator variables for day of week, temp_{ij} and $\overline{\mathrm{temp}}^{(3)}_{ij}$ are respectively the current day's

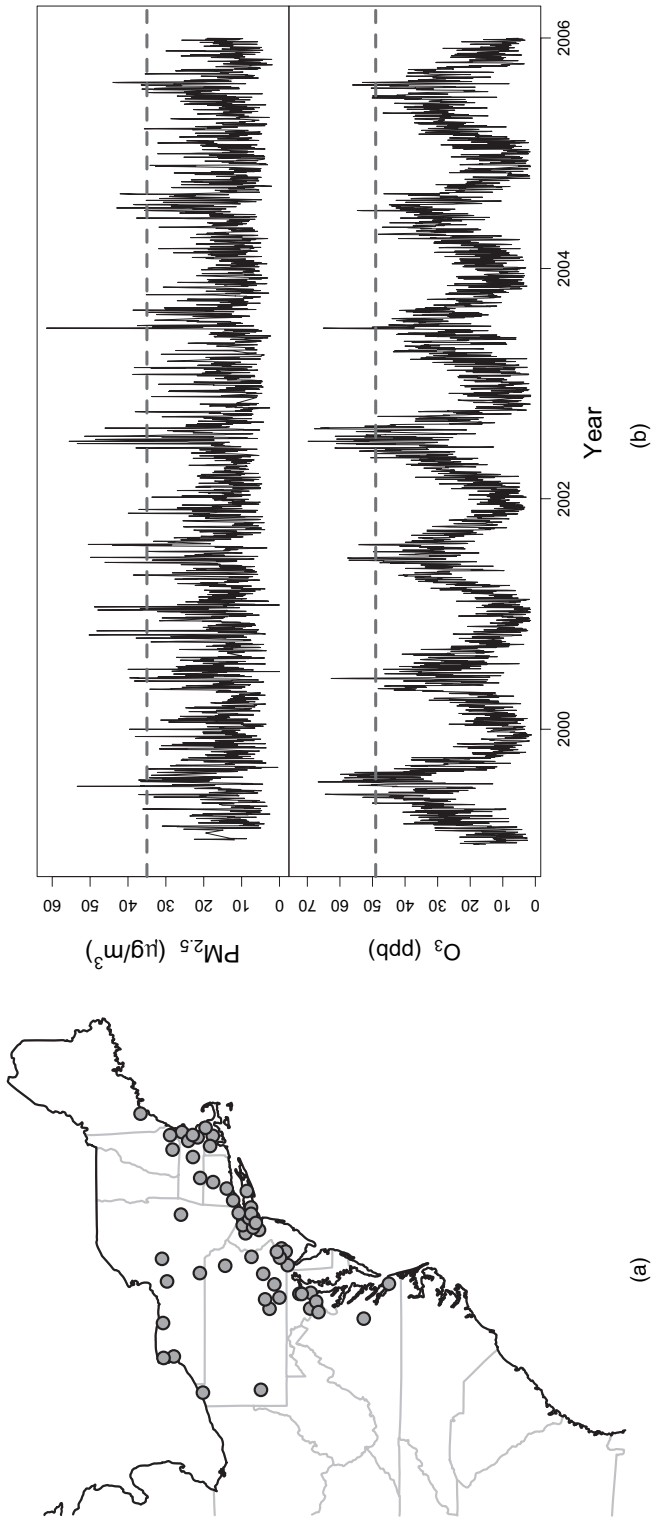

Fig. 1. (a) Map of 51 north-eastern US counties used for a multisite time series study of the association between joint exposure to $PM_{2.5}$ and O_3 and hospitalization for CVDs and (b) daily time series of $PM_{2.5}$ and O_3 for the District of Columbia for the period 1995–2005: – – –, daily national standard for each pollutant

and the average of the previous 3 days' average temperature and dptp_{ij} and $\overline{\text{dptp}}_{ij}^{(3)}$ are respectively the current day's and the average of the previous 3 days' average dewpoint temperature. The smooth function of calendar time $\text{ns}(j; 7\,\text{DF/year}, \gamma_{i7})$ accounts for seasonality and longer-term, time varying trends in hospital admissions.

This within-county model extends those developed to study $PM_{2.5}$ and O_3 individually (Dominici *et al.*, 2006; Bell *et al.*, 2004) by allowing for non-linear associations of each of the pollutants and their interaction. In particular, the choice of covariates and DFs in the smooth functions are based on those used by Dominici *et al.* (2006). Previous studies have assessed the sensitivity of health effect estimates from single-pollutant models to adjustment for temperature and the smooth function of calendar time, finding that results were robust across specifications of the confounder model (Peng *et al.*, 2006; Welty and Zeger, 2005).

To place model (1) within the more general context of HMs for two-level clustered data, we introduce some notation. Let $\mathbf{b}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2})$ be the vector of random effects for the exposure–response surface characterizing the relationship between joint exposure to O_3 and fine particulate matter and the health outcome. Let $\gamma_i = (\gamma_{i0}, \gamma_{i1}, \ldots, \gamma_{i7})$ be the vector of random effects describing the association between the confounders and the health outcome, and define $\boldsymbol{\beta}_i = (\mathbf{b}_i, \gamma_i)$. Note that these random effects are introduced to model variation across counties, not as a random-effects parameterization of penalized splines (the number of DFs in the spline terms is fixed). Let $\mathbf{x}_{ij}$ denote the full vector of covariate data for day j in county (cluster) i, and let $\mathbf{x}_{ij}^{\mathbf{b}}$ denote the 15-dimensional subvector of $\mathbf{x}_{ij}$ that is the concatenation of the basis terms for the main effects and interactions of the spline bases for ozone and fine particulate matter, $\text{ns}(PM_{2.5_{ij}}; 3\,\text{DF}, \mathbf{b}_{i1})\,\text{ns}(O_{3_{ij}}; 3\,\text{DF}, \mathbf{b}_{i2})$.

We next define a variable that identifies whether the daily levels of either $PM_{2.5}$ and/or O_3 are above or below their corresponding 24-h NAAQSs,

$$\text{NAAQS}_{ij} = \begin{cases} A & \text{if } PM_{2.5} > 35\,\mu\text{g}\,\text{m}^{-3} \text{ and } O_3 > 0.049 \text{ ppm}, \\ B & \text{if } PM_{2.5} > 35\,\mu\text{g}\,\text{m}^{-3} \text{ and } O_3 \leqslant 0.049 \text{ ppm}, \\ C & \text{if } PM_{2.5} \leqslant 35\,\mu\text{g}\,\text{m}^{-3} \text{ and } O_3 > 0.049 \text{ ppm}, \\ D & \text{if } PM_{2.5} \leqslant 35\,\mu\text{g}\,\text{m}^{-3} \text{ and } O_3 \leqslant 0.049 \text{ ppm}. \end{cases}$$

The values 35 $\mu\text{g}\,\text{m}^{-3}$ and 0.049 ppm were derived from the NAAQSs, which are defined in appendix A of the on-line supplementary materials.

We define θ_i to be the logarithm of the expected number of CVD admissions on days when both $PM_{2.5}$ and O_3 are above their respective national standards divided by the expected number of CVD admissions on days when both pollutants are lower than their national standards, adjusted for the potential confounding variables:

$$\theta_i := h(\boldsymbol{\beta}_i; \mathbf{x}_i) = \log \left\{ \frac{(1/N_{iA}) \sum_{j:\text{NAAQS}_{ij}=A} \exp(\mathbf{b}_i' \mathbf{x}_{ij}^{\mathbf{b}})}{(1/N_{iD}) \sum_{j:\text{NAAQS}_{ij}=D} \exp(\mathbf{b}_i' \mathbf{x}_{ij}^{\mathbf{b}})} \right\}. \tag{2}$$

Here N_{iA} and N_{iD} are the number of days when both pollutants are respectively above or below their respective national standards in county i during the study period 1999–2005. Derivation of the formulation for the parameter of interest is in appendix B of the on-line supplementary materials. Other definitions of θ_i that may be of interest, such as the logarithm of the expected number of CVD admissions on days when only the level of $PM_{2.5}$ (or when only O_3) is above its national standard divided by the expected number of CVD admissions on days when both pollutants are lower than their national standards could be defined similarly and the same methods (which are described below) could be straightforwardly applied.

4. Reduced hierarchical model

Rather than specify a full HM on the large number of random effects $\boldsymbol{\beta}_i$, we define a reduced HM directly on the parameter of interest θ_i:

$$\begin{aligned} \mathbf{y}_i|\theta_i &\sim L_i(\theta_i), \qquad \text{independent}, i=1,\ldots,I, \\ \theta_i|\boldsymbol{\alpha} &\sim \mathrm{RE}(\theta_i|\boldsymbol{\alpha}), \qquad \text{independent}, i=1,\ldots,I. \end{aligned} \tag{3}$$

Here $L_i(\theta_i)$ denotes a likelihood function (which is detailed below) and $\mathrm{RE}(\theta_i|\boldsymbol{\alpha})$ denotes an arbitrary random-effect distribution. Note that the likelihood function in general depends on the vector of outcome data from the ith cluster $\mathbf{y}_i$ and on the set of covariate data $\mathbf{x}_i$, though we suppress this dependence in our notation. To conduct inference in the Bayesian framework, a prior distribution is placed on $\boldsymbol{\alpha}$.

The reduced HM may be further generalized by allowing the random-effect distribution $\mathrm{RE}(\theta_i|\boldsymbol{\alpha})$ to depend on cluster level covariates $\mathbf{z}_i$, to study potential effect modification. In particular, for the second-stage model we assume that $\theta_i = \alpha_{0i} + \boldsymbol{\alpha}_1'\mathbf{z}_i$ and place the random-effect distribution on the α_{0i}. The second-level model may also be extended to allow the θ_i to be spatially correlated across clusters.

4.1. Integrated likelihood

In the general setting where the parameter of interest θ_i is a complicated function of the level 1 parameters $\boldsymbol{\beta}_i$ as in expression (2), we propose to use an integrated likelihood for $L_i(\theta_i)$. For notational simplicity the cluster-specific subscript i is suppressed in what follows. An integrated likelihood for the ith cluster may be expressed as

$$f_{\mathbf{y}|\theta}(\mathbf{y}|\theta) \propto f_{\theta|\mathbf{Y}}(\theta|\mathbf{y})/\pi_\theta(\theta), \tag{4}$$

where $\pi_\theta(\theta)$ is the prior distribution for θ and $f_{\theta|\mathbf{Y}}$ is the corresponding posterior distribution of θ based on the data from only that cluster. Note that, in the special case where the cluster-specific parameters $\boldsymbol{\beta}$ can be reparameterized as $(\theta, \boldsymbol{\lambda})$, this expression can be rewritten as

$$f_{\mathbf{y}|\theta}(\mathbf{y}|\theta) = \int f_{\mathbf{y}|\theta,\boldsymbol{\lambda}}(\mathbf{y}|\theta,\boldsymbol{\lambda})\, \pi_{\boldsymbol{\lambda}|\theta}(\boldsymbol{\lambda}|\theta)\, \mathrm{d}\boldsymbol{\lambda},$$

where $f_{\mathbf{y}|\theta,\boldsymbol{\lambda}}$ is the joint likelihood, and $\pi_{\boldsymbol{\lambda}|\theta}$ is the prior density of $\boldsymbol{\lambda}$ given θ (Berger *et al.*, 1999).

When such a reparameterization of $\boldsymbol{\beta}$ is not available or when $f_{\mathbf{y}|\theta}(\mathbf{y}|\theta)$ is not available in closed form, we propose a simulation approach to approximate expression (4) as follows.

> *Step 1*: assign a prior distribution to the vector $\boldsymbol{\beta}$ of level 1 parameters, such that the induced prior distribution $\pi_\theta(\theta)$ on $\theta = h(\boldsymbol{\beta}; \mathbf{x})$ is diffusely spread out over the range of plausible values for θ. Simulate R prior samples from $\pi_\theta(\theta)$.
> *Step 2*: fit a within-cluster model to generate R samples $\boldsymbol{\beta}^{(r)}$ from the posterior $f_{\boldsymbol{\beta}|\mathbf{y}}(\boldsymbol{\beta}|\mathbf{y})$.
> *Step 3*: obtain the posterior samples $\theta^{(r)} = h(\boldsymbol{\beta}^{(r)}; \mathbf{x})$.
> *Step 4*: select a grid of points $\{\theta_k\}$ covering the range of θ and apply a Gaussian kernel smoother to estimate both $f_{\theta|\mathbf{y}}(\theta|\mathbf{y})$ and $\pi_\theta(\theta)$ on this grid.

We repeat this process for each cluster i to obtain approximations $\hat{f}_{\mathbf{y}_i|\theta_i}(\mathbf{y}_i|\theta_i)$, $i=1,\ldots,I$. Note that the choice of prior distribution for $\boldsymbol{\beta}$ in step 1 will depend on the form of the function h. Also note that although this procedure requires drawing from the posterior $f_{\boldsymbol{\beta}_i|\mathbf{y}_i}(\boldsymbol{\beta}_i|\mathbf{y}_i)$, since this is done within each cluster independently, the sampling is greatly simplified compared with fitting the full HM where the $\boldsymbol{\beta}_i$ are correlated across clusters (i.e. sampling from

$f_{\beta_1,\ldots,\beta_I|\mathbf{y}_1,\ldots,\mathbf{y}_I}(\beta_1,\ldots,\beta_I|\mathbf{y}_1,\ldots,\mathbf{y}_I)$). In addition, since this step is performed a single time before fitting the reduced HM, estimating the parameters of the reduced HM remains fast. Further details of our implementation are in appendix C of the on-line supplementary materials.

4.2. Dirichlet process mixture model for RE($\theta_i|\alpha$)

To allow for flexible specification of the random-effect distribution we propose to use a Dirichlet process mixture model for $\mathrm{RE}(\theta_i|\boldsymbol{\alpha})$. The Dirichlet process mixture model (Ferguson, 1973; Neal, 2000) can be expressed as the limit as the number of components $K \to \infty$ of the mixture model

$$\begin{aligned}
\theta_i|c_i, \boldsymbol{\phi} &\sim F(\theta_i|\boldsymbol{\phi}_{c_i}), && \text{independent, } i = 1,\ldots,I,\\
c_i|\mathbf{p} &\sim \text{discrete}(c_i|p_1,\ldots,p_K), && \text{independent, } i = 1,\ldots,I,\\
\phi_c &\sim G_0 \text{ for any } c,\\
\mathbf{p} &\sim \text{Dirichlet}(\delta/K,\ldots,\delta/K),
\end{aligned}$$

where discrete$(c_i|p_1,\ldots,p_K)$ corresponds to the probability mass function $\mathbb{P}(c_i = k) = p_k$ $(k = 1,\ldots,K)$ and δ/K is the concentration parameter written so that it approaches 0 as $K \to \infty$. Here we consider a normal mixture so that $F(\cdot|\boldsymbol{\phi}_c) = N(\cdot|\mu_c, \tau_c)$, and we select the conjugate prior so that $G_0 = \text{normal gamma}(\lambda, \gamma, a, b)$, i.e. $\tau_c \sim \text{gamma}(\tau|a, b)$ and $\mu_c|\tau_c \sim N(\lambda, \gamma\tau_c)$.

4.3. Computational details

The reduced HM (3) may be fitted by using MCMC methods (Metropolis *et al.*, 1953; Gilks *et al.*, 1995) to generate samples from the posterior distribution of the unknown parameters

$$\mathbb{P}(\theta_1,\ldots,\theta_I,\boldsymbol{\alpha}|\mathbf{y}_1,\ldots,\mathbf{y}_I) \propto \pi(\boldsymbol{\alpha}) \prod_{i=1}^{I} \{\mathrm{RE}(\theta_i|\boldsymbol{\alpha})\, L_i(\theta_i)\},$$

where $\pi(\boldsymbol{\alpha})$ denotes the prior distribution on the vector of parameters of the random-effect distribution. At each iteration of the MCMC algorithm, a sample is drawn from the full conditional

$$f_c(\theta_i) \propto \mathrm{RE}(\theta_i|\boldsymbol{\alpha})\, L_i(\theta_i) \tag{5}$$

for each cluster i. When the integrated likelihood has been estimated by using the approach from Section 4.1, we replace $L_i(\theta_i)$ in equation (5) by $\hat{f}_{\mathbf{y}_i|\theta_i}(\mathbf{y}_i|\theta_i)$. Since $f_c(\theta_i)$ is not a known distribution, we sample from it by applying a Metropolis–Hastings step. In the Metropolis–Hastings step, we need to evaluate the likelihood $\hat{f}_{\mathbf{y}_i|\theta_i}$ at an arbitrary point θ. We do this by selecting the grid point θ_k that is closest to θ and evaluating the likelihood at that grid point.

For generating posterior samples of $\boldsymbol{\alpha}$ when $\mathrm{RE}(\theta_i|\boldsymbol{\alpha})$ is the Dirichlet process mixture model that was defined in Section 4.2, we adapt an MCMC sampling algorithm described by Neal (2000). Details are in appendix C of the on-line supplementary materials.

5. Simulation study

There are instances for which the reduced HM may be preferred to the full HM owing to practical considerations such as its simplified implementation and the ease with which prior information may be incorporated directly on the parameter of interest. However, a more thorough understanding of situations when the reduced HM works well is needed. In this section we conduct simulation studies to compare performance of the reduced HM with the full HM across a range of scenarios.

We base our studies on data from a meta-analysis of 41 randomized trials of a treatment for stomach ulcers, provided by Efron (1996). Rather than use the multipollutant case-study as a basis for simulation studies, a meta-analysis example is used to highlight the broad utility of the reduced HM methodology across diverse applications. In addition, even in the simpler context of this application (a two-dimensional vector of random effects β_i) for which a full HM may be straightforwardly implemented, the relative performance of the reduced HM to the full HM is not well understood and, as we shall see, the full HM may not always be the optimal choice even in the low dimensional case.

The data from the ith trial are $\{\mathbf{y}_i = (y_{i0}, y_{i1}), \mathbf{x}_i = (n_{i0}, n_{i1})\}$, where y_{i0} and y_{i1} are the number of occurrences of ulcers for the control and treatment groups, and n_{i0} and n_{i1} are the number of subjects in the control and treatment groups respectively. Let $\mathbf{p}_i = (p_{i0}, p_{i1})$ be the vector of probabilities of the occurrence of ulcers in the control and treatment groups. The distribution of the data from experiment (cluster) i is assumed to be

$$\mathbb{P}_i(\mathbf{y}_i|\mathbf{x}_i;\mathbf{p}_i) = \binom{n_{i1}}{y_{i1}} p_{i1}^{y_{i1}}(1-p_{i1})^{n_{i1}-y_{i1}} \binom{n_{i0}}{y_{i0}} p_{i0}^{y_{i0}}(1-p_{i0})^{n_{i0}-y_{i0}},$$

and the parameter of interest is the log-odds ratio

$$\theta_i = h(\mathbf{p}_i) = \log\left\{\frac{p_{i1}/(1-p_{i1})}{p_{i0}/(1-p_{i0})}\right\}. \tag{6}$$

In this example, a full HM would require the specification of a random-effect distribution for $\mathbf{p}_i = (p_{i1}, p_{i0})$. Alternatively, a commonly used specification first defines a one-to-one transformation of the $\mathbf{p}_i$ into $\mathbb{R}^2$ through the logit link and assumes a bivariate normal distribution for the random effects:

$$\begin{aligned} y_{ki}|p_{ik} &\sim \text{Binom}(n_{ik}, p_{ik}) \qquad \text{for } k = 0, 1, \\ \text{logit}(p_{ik}) &= \beta_{i0} + \beta_{i1}\mathbf{I}(k=1), \\ (\beta_{i0}, \beta_{i1})' &\sim N\{(\beta_0^*, \beta_1^*)', \Sigma\}. \end{aligned} \tag{7}$$

For a reduced HM, we first summarize the information that is contained in experiment i about the log-odds ratio θ_i through a likelihood function, and we then specify a random-effect distribution directly on the θ_i. For this problem, a conditional likelihood for θ_i is available in closed form. By conditioning on the margins of the 2×2 table for each experiment, the conditional likelihood may be expressed as

$$L_i^{\text{C}}(\theta_i) = \frac{\binom{n_{i0}}{y_{i0}}\binom{n_{i1}}{y_{i1}}\exp(\theta_i y_{i1})}{\sum_{u=0}^{\min(n_{i1}, y_{i0}+y_{i1})}\binom{n_{i0}}{u}\binom{n_{i1}}{y_{i1}+y_{i0}-u}\exp(\theta_i u)}. \tag{8}$$

We may then use $L_i^{\text{C}}(\theta_i)$ for the likelihood function in the reduced HM (3). Computing integrated likelihoods for each of the randomized trials in the ulcer data set (Efron, 1996), we found them to be generally quite similar to the corresponding conditional likelihoods, and so only the conditional likelihoods were considered in the simulation study.

We simulated data under four data-generating mechanisms, and we estimated model parameters under four HM formulations. We next describe each of the hierarchical modelling approaches that were used to fit the data, after which we detail the four data-generating models.

5.1. Hierarchical models

We fitted each simulated data set by using four approaches: a full HM assuming the logistic model (7) with a normal random-effect distribution on the $\boldsymbol{\beta}_i$ (model FHM); a reduced HM using the conditional likelihood $L_i^{\mathrm{C}}(\theta_i)$ from equation (8) with a normal random-effect distribution on the θ_i (model RHM-L-N); a reduced HM using the conditional likelihood $L_i^{\mathrm{C}}(\theta_i)$ from equation (8) with a flexible random-effect distribution on the θ_i (model RHM-L-DP); a reduced HM using a normal approximation to the likelihood with a normal random-effect distribution on the θ_i (model RHM-N-N). For the flexible random-effect distribution, we considered the Dirichlet process normal mixture model that was described in Section 4.2. For each approach, we estimated the cluster-specific log-odds ratios θ_i as well as the overall log-odds ratio $\theta^* = \mathbb{E}(\theta_i)$, where the expectation is taken over all of the clinical trials that were included in the analysis. Additionally, we obtained 95% posterior intervals for the overall (θ^*) and cluster-specific (θ_i) parameters. Details of estimation for each of the four models are in appendix D of the on-line supplementary materials.

5.2. Data-generating models

We considered four data-generating models. We always assumed $y_{i0} \sim \mathrm{Binom}(n_{i0}, p_{i0})$ and $y_{i1} \sim \mathrm{Binom}(n_{i1}, p_{i1})$, and we selected different models for generating p_{i0} and p_{i1} ($i = 1, \ldots, I$). Note that each model for generating p_{i0} and p_{i1} induces a distribution on the log-odds ratio θ_i through expression (6). Thus, each time that we generated a data set, we obtained I values of the cluster-specific, true log-odds ratios θ_i (one for each cluster i). The models were selected to distinguish between scenarios where the full HM is expected to outperform the reduced HM and vice versa. Fig. 2 shows, for each of the four data-generating models, the distribution of the (p_{i0}, p_{i1}), along with the corresponding distributions of the $(\beta_{0i}, \beta_{1i}) = (\log\{p_{i0}/(1-p_{i0})\}, \theta_i)$ and the log-odds ratios θ_i.

In each case, we set $n_{i0} = n_{i1} = n$, and we considered $n = 100$ for $I = 100, 50, 25$. These parameter values were selected to correspond to a large within-cluster sample size for either a large, moderate or small number of clusters.

5.2.1. Model 1—bivariate normal

We generated data from

$$(\beta_{0i}, \beta_{1i})' \sim N\{(\beta_0^*, \beta_1^*)', \Sigma\},$$
$$\mathrm{logit}(p_{ki}) = \beta_{0i} + \beta_{1i}\,\mathbf{I}(k = 1),$$

where $(\beta_0^*, \beta_1^*) = (-0.2, -1.3)$, and we considered two different values for Σ,

$$\Sigma_{\mathrm{a}} = \begin{pmatrix} 0.9 & 0 \\ 0 & 1.1 \end{pmatrix} \quad \text{and} \quad \Sigma_{\mathrm{b}} = \begin{pmatrix} 0.9 & 0.5 \\ 0.5 & 1.1 \end{pmatrix}$$

(see scenarios 1(a) and 1(b) in Fig. 2). These parameter values were selected to be the same order of magnitude as those from the ulcer data set. Since this model fully specifies a normal random-effect distribution on the $\boldsymbol{\beta}_i$, particularly in scenario 1(b) where a moderate correlation between the random effects is assumed, we expected it to favour the full HM (7).

5.2.2. Model 2—uniform–beta

We generated $p_{i0} \sim \mathrm{uniform}(0.1, 0.6)$ and $p_{i1}|p_{i0} \sim \mathrm{beta}(m = p_{i0} + 0.3, \phi)$, where the beta distribution is parameterized by its mean m and variance ϕ. We considered two values for ϕ, namely

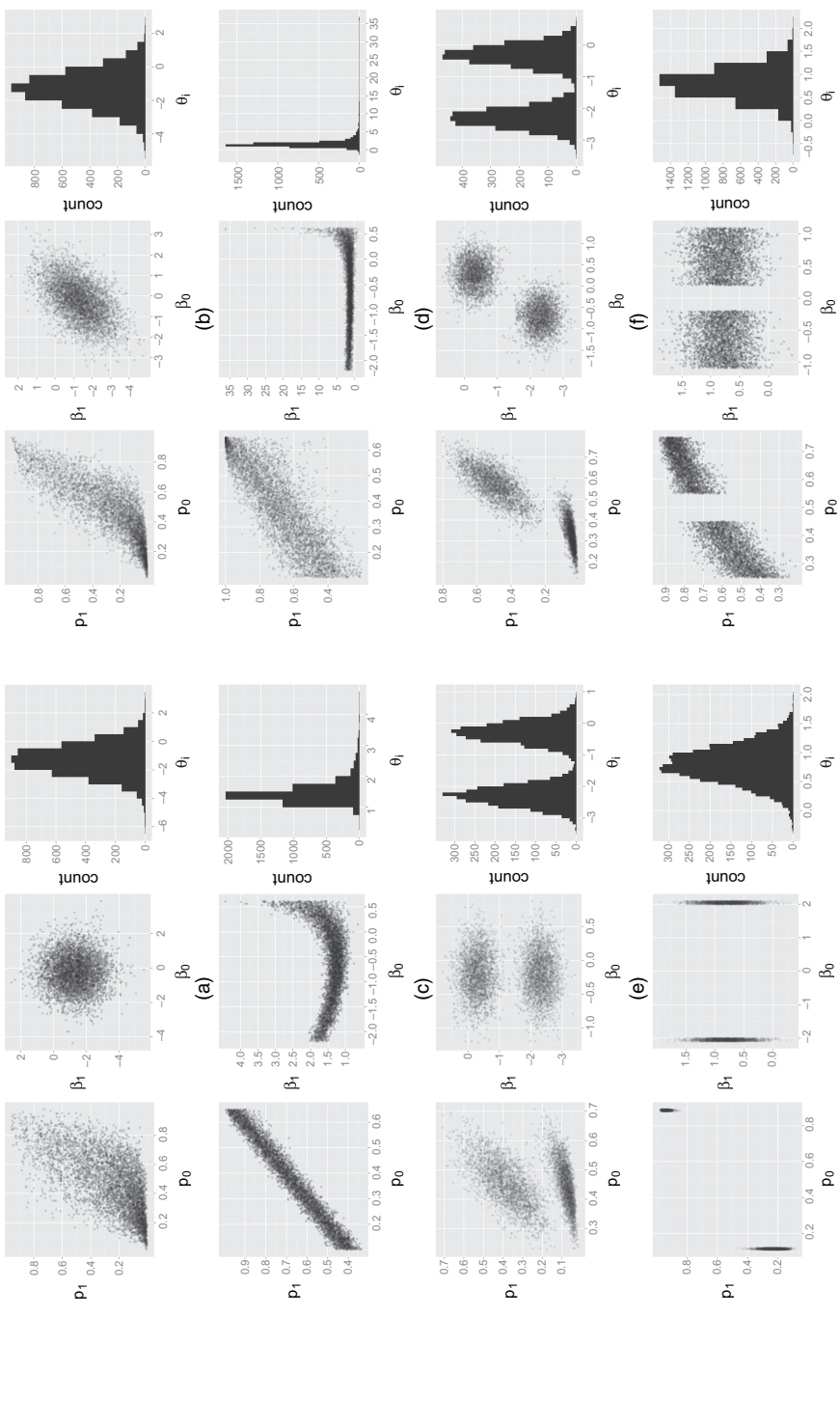

Fig. 2. Plots of simulated data under each scenario for four data-generating models (for each scenario, 5000 data points (p_{i0}, p_{i1}) are plotted, as well as the corresponding points (β_{i0}, β_{i1}) under the transformation $\text{logit}(p_{ik}) = \beta_{i0} + \beta_{i1}\mathbf{I}(k = 1)$), and histograms of the corresponding log-odds ratios θ_i: (a) model 1, scenario 1(a); (b) model 1, scenario 1(b); (c) model 2, scenario 2(a); (d) model 2, scenario 2(b); (e) model 3, scenario 3(a); (f) model 3, scenario 3(b); (g) model 4; scenario 4(a); (h) model 4, scenario 4(b)

$\phi_a = 0.001$ and $\phi_b = 0.01$. Since this model is not based on either the full or reduced HM *a priori* we did not expect it to favour either of these two approaches (see scenarios 2(a) and 2(b) in Fig. 2).

5.2.3. *Model 3—normal mixture*

We generated (p_{i0}, p_{i1}) by

$$(\beta_{0i}, \beta_{1i})' \sim \alpha\, N(\boldsymbol{\beta}^* - \boldsymbol{\nu}, \Sigma) + (1-\alpha)\, N(\boldsymbol{\beta}^* + \boldsymbol{\nu}, \Sigma),$$
$$\text{logit}(p_{ki}) = \beta_{0i} + \beta_{1i}\, \mathbf{I}(k=1),$$

where we fixed $\boldsymbol{\beta}^* = (-0.2, 1.3)$, $\alpha = 0.5$ and $\Sigma = \text{diag}(0.1, 0.1)$. We considered two values for $\boldsymbol{\nu}$, namely $\boldsymbol{\nu}'_a = (0, 1)$ and $\boldsymbol{\nu}'_b = (0.5, 1)$. This data-generating model was selected because the random-effect distribution will be misspecified for both the full and the reduced HM (since $\theta_i = \beta_{i1}$), when a normal random-effect distribution is assumed; thus, we expected neither approach to perform particularly well (see scenarios 3(a) and 3(b) in Fig. 2).

5.2.4. *Model 4—normal–θ_i*

Finally, we generated data by first simulating values for the log-odds ratios θ_i and for the log-odds $\lambda_i = \log\{p_{0i}/(1-p_{0i})\}$, which induces a distribution on the

$$(p_{0i}, p_{1i}) = \left(\frac{\exp(\lambda_i)}{1+\exp(\lambda_i)}, \frac{\exp(\lambda_i+\theta_i)}{1+\exp(\lambda_i+\theta_i)} \right).$$

In particular, we simulated $\theta_i \sim N(\mu, \sigma^2)$ and $\lambda_i \sim 0.5\, U(-u_2, -u_1) + 0.5\, U(u_1, u_2)$, where we fixed $\mu = 0.8$ and $\sigma^2 = 10$. We considered two scenarios for u_1 and u_2, namely $(u_{1a}, u_{2a}) = (2, 2.1)$ and $(u_{1b}, u_{2b}) = (0.2, 1.1)$. This model was chosen because it was expected to favour the reduced HM over the full HM, since the normal random-effect distribution on the $(\beta_{i0}, \beta_{i1})'$ for the full HM will be misspecified, whereas the random-effect distribution for θ_i in the reduced HM will be correctly specified (see scenarios 4(a) and 4(b) in Fig. 2).

5.3. *Results*

We evaluated the relative performance of the four modelling approaches (FHM, RHM-L-N, RHM-L-DP and RHM-N-N) in estimating both the cluster-specific (θ_i) and overall (θ^*) log-odds ratios. Because disparity in performance across methods was attenuated for the smaller values for the numbers of clusters, in this section we focus our discussion on results for $I = 100$ (Table 1). Results for the cases $I = 25$ and $I = 50$ are in Tables S1 and S2 of the on-line supplementary materials.

The main disparity in performance across the reduced HM (models RHM-L-N and RHM-L-DP) and full HM approaches occurred for estimation of the cluster-specific parameters θ_i; methods (except RHM-N-N) performed comparably for estimating the overall θ^*. The two situations where model FHM yielded similar or slightly better cluster-specific estimates than the reduced HM were those for which the data-generating model implied considerable correlation between β_{0i} and β_{1i}, which could be captured to varying degrees by the bivariate normal random-effect distribution on the $\boldsymbol{\beta}_i$. This occurred for data-generating models 1(b) and 3(b), which had correlations of about 0.5 and 0.8 respectively (see Fig. 2). Because nuisance parameters are eliminated before pooling, the reduced HMs do not take advantage of this correlation structure. For the other scenarios, the reduced HM generally performed comparably with or better than FHM. Comparing the reduced HM with different random-effect distributions, we found that RHM-L-DP performed just as well as or only slightly worse than RHM-L-N when

Table 1. Simulation results for the cluster-specific log-odds ratios θ_i: squared error loss $\Sigma_{i=1}^{I}(\tilde{\theta}_i - \theta_i)^2$ for the posterior mean estimates $\tilde{\theta}_i$ and coverage of 95% posterior intervals†

Model	*Simulation*	*Cluster* θ_i		*Overall* θ^*			
		Squared error loss	*Coverage*	*Bias*	*Standard deviation*	*Root-mean-squared error*	*Coverage*
Model 1, bivariate normal	*1(a)*‡			$\theta^* = -1.3$			
	FHM	14.4	0.95	0.00	0.11	0.11	0.94
	RHM-L-N	14.8	0.95	0.02	0.11	0.11	0.95
	RHM-L-DP	14.8	0.95	0.03	0.11	0.11	0.94
	RHM-N-N	18.0	0.94	0.09	0.10	0.14	0.89
	1(b)‡			$\theta^* = -1.3$			
	FHM	14.9	0.95	−0.01	0.12	0.12	0.94
	RHM-L-N	18.9	0.94	0.04	0.11	0.12	0.93
	RHM-L-DP	18.9	0.94	0.06	0.11	0.12	0.92
	RHM-N-N	27.5	0.92	0.14	0.10	0.17	0.74
Model 2, uniform–beta	*2(a)*			$\theta^* = 1.46$			
	FHM	7.5	0.88	−0.03	0.04	0.05	0.87
	RHM-L-N	8.0	0.91	−0.03	0.04	0.05	0.90
	RHM-L-DP	7.5	0.95	−0.04	0.04	0.06	0.90
	RHM-N-N	9.6	0.89	−0.07	0.04	0.08	0.66
	2(b)			$\theta^* = 1.67$			
	FHM	99.5	0.90	−0.14	0.08	0.16	0.55
	RHM-L-N	104.2	0.91	−0.13	0.08	0.16	0.56
	RHM-L-DP	96.2	0.92	−0.14	0.09	0.17	0.57
	RHM-N-N	137.6	0.89	−0.23	0.07	0.24	0.11
Model 3, normal mixture	*3(a)*			$\theta^* = -1.3$			
	FHM	10.6	0.95	−0.01	0.11	0.11	0.97
	RHM-L-N	11.6	0.95	0.00	0.11	0.11	0.97
	RHM-L-DP	9.8	0.96	0.02	0.10	0.11	1.00
	RHM-N-N	11.6	0.95	0.06	0.10	0.12	0.94
	3(b)			$\theta^* = -1.3$			
	FHM	10.1	0.95	−0.02	0.12	0.13	0.93
	RHM-L-N	13.5	0.95	0.02	0.12	0.12	0.94
	RHM-L-DP	12.1	0.96	0.03	0.12	0.12	0.99
	RHM-N-N	15.0	0.94	0.12	0.11	0.17	0.80
Model 4, normal–θ_i	*4(a)*			$\theta^* = 0.8$			
	FHM	7.9	0.84	0.00	0.06	0.06	0.85
	RHM-L-N	7.2	0.93	0.01	0.06	0.06	0.94
	RHM-L-DP	7.3	0.97	−0.01	0.06	0.06	0.96
	RHM-N-N	7.5	0.90	−0.05	0.05	0.07	0.86
	4(b)			$\theta^* = 0.8$			
	FHM	5.1	0.93	0.00	0.05	0.05	0.93
	RHM-L-N	5.2	0.94	0.00	0.05	0.05	0.94
	RHM-L-DP	5.2	0.96	−0.01	0.04	0.05	0.94
	RHM-N-N	5.2	0.94	−0.02	0.04	0.05	0.93

†Results for the mean log-odds ratio θ^*: bias, standard deviation and root-mean-squared error of the posterior mean estimates $\tilde{\theta}^*$ and coverage of 95% posterior intervals. Methods compared are the full HM (FHM), the reduced HM with conditional likelihood and normal random-effect distribution (RHM-L-N), the reduced HM with normal approximation to the likelihood and normal random-effect distribution (RHM-N-N) and the reduced HM with conditional likelihood and Dirichlet process normal mixture for the random-effect distribution (RHM-L-DP).

‡For scenarios 1(a) and 1(b), the summary statistics for model RHM-L-DP are based on 999 and 998 simulation repetitions respectively. The other repetitions were excluded because the MCMC algorithm did not converge within the maximum number of iterations.

the true distribution was normal (models 1(a) and 1(b) and 4(a) and 4(b)) but performed moderately better when the true random-effect distribution was non-normal (models 2(a) and 2(b) and 3(a) and 3(b)).

Across simulation scenarios we generally found that the model using the normal approximation to the likelihood (model RHM-N-N), although most efficient computationally, was not competitive with the other approaches. For estimating θ_i, model RHM-N-N either performed comparably (scenarios 2(a), 3(a) and 4(a) and 4(b)), or moderately worse (scenarios 1(a) and 1(b), 2(b) and 3(b)) than the other approaches. For estimating the overall θ^*, RHM-N-N generally had larger root-mean-squared error rMSE and coverage markedly lower than the nominal rate (exceptions are scenarios 3(a) and 4(c)). One reason for the poor performance of RHM-N-N is that the normal approximation to the likelihood does not provide a good approximation in this application, particularly when y_{i1} or y_{i0} is equal to 0 or n (which occurs most frequently under models 1(a) and 1(b) and 3(a), scenarios where RHM-N-N performs worst). In addition, we note that, under scenario 2(b), none of the approaches performed particularly well for estimating the mean (θ^*) of the highly skewed random-effect distribution for θ_i.

5.4. Conclusions

Our simulation studies were designed to assess the relative performance of the reduced HM to the full HM across different scenarios of misspecification of the random-effect distribution. We found that large correlation in the random-effects β_i generally led to slightly improved estimation of the cluster-specific θ_i by the full HM compared with the reduced HM. However, in other scenarios, namely those for which the random-effect distribution for the full HM was misspecified, the reduced HM achieved superior performance. In addition, for estimating the overall θ^* we found performance to be very similar across methods. Overall, in our simulation studies the reduced HM performed nearly as well as the full HM, and even performed better in some cases.

6. Application

We applied the reduced HM to our multisite time series study of 51 urban counties in north-eastern USA for the period 1999–2005. Our goal was to estimate the county-specific and overall log-relative-risks of emergency cardio-vascular hospital admissions associated with levels of $PM_{2.5}$ and O_3 above their national standards.

We considered three types of reduced HM. The first uses a normal approximation to the likelihood at the first stage and a normal random-effect distribution at the second stage (model RHM-N-N). The second uses an integrated likelihood at the first stage and a normal random-effect distribution at the second stage (model RHM-L-N). The third uses an integrated likelihood at the first stage and a Dirichlet process normal mixture for the random-effect distribution (model RHM-L-DP). The parameter of interest θ_i, defined in expression (2), is the log-relative-risk of cardio-vascular admissions when $PM_{2.5}$ and O_3 are both above their national standards compared with when both are below their standards. For each reduced HM we assumed little prior information, by incorporating diffuse priors on the overall θ^*. We first fitted each reduced HM without including any second-level covariates. We subsequently considered inclusion, at the second stage, of a county-specific measure of the average level of nitrogen dioxide (NO_2) during the study period to demonstrate how reduced HMs may be used to identify effect modification. Long-term average NO_2 levels may be an important effect modifier because they are a proxy for traffic exposure. This was done by assuming, at the second level, that $\theta_i = \alpha_{0i} + \alpha_1 z_i$, where z_i is the long-term average NO_2 level for the ith county, and placing each of the normal

($\alpha_{0i} \sim N(\alpha_0^*, \tau^2)$) and flexible (Section 4.2) random-effect distributions on the α_{0i}. Details of the implementations for each reduced HM are in appendix C of the on-line supplementary materials.

Before fitting the reduced HM by using the integrated likelihood (models RHM-L-N and RHM-L-DP), we evaluated the performance of the integrated likelihood in the air pollution context through simulation study (detailed in appendix E of the on-line supplementary materials). Briefly, we considered a model based on our air pollution and health outcome data for which the integrated likelihood may be written in closed form. We simulated data under this model, applied our approach to estimate the integrated likelihood (described in Section 4.1) and compared our estimated integrated likelihood with the true integrated likelihood, finding that the estimate closely matched the truth.

Fig. 3 shows the posterior mean estimates and 95% posterior intervals for the overall θ^* and for the cluster-specific θ_i obtained under each reduced HM. We found that on average, across all counties, there was an increase in CVD admissions on days when both O_3 and fine particulate matter were above their national standards compared with days when both pollutants were below their national standards. In particular, we estimated that the overall log-relative-risk of CVD admissions associated with levels of O_3 and $PM_{2.5}$ both above their national standards (θ^*) was 0.024 (95% posterior interval −0.004–0.053) for model RHM-N-N, 0.027 (interval −0.007–0.061) for model RHM-L-N and 0.029 (interval −0.014–0.071) for model RHM-L-DP. A log-relative-risk of 0.024 corresponds (approximately) to an increase of 2.4% in cardio-vascular hospital admissions on days when both O_3 and $PM_{2.5}$ are above their standards compared with days when both pollutants are below their standards. We also found variability across counties in the estimate of the cluster-specific effects θ_i. For most counties, θ_i was estimated to be positive, though for each county the posterior interval covered zero. The random-effect estimates exhibited the largest shrinkage for model RHM-N-N, followed by RHM-L-N, with the RHM-L-DP estimates remaining furthest from the overall regional estimate.

Fig. 4 shows the posterior mean estimates of the location-specific θ_i from the reduced HM including average NO_2 level as a covariate at the second stage, plotted against the location's long-term average NO_2 level. The positive slopes (α_1) suggest that the risk of cardio-vascular admissions that are associated with daily levels of O_3 and $PM_{2.5}$ that are greater than their national standards is higher in locations with greater NO_2 levels and lower in locations with lower NO_2 levels, though the estimates were not statistically significant. More precisely, we estimated that an interquartile range increase in long-term average NO_2 level is associated with a percentage increase in the relative risk of cardio-vascular hospital admissions associated with O_3 and $PM_{2.5}$ both above their national standards of 1.2% (95% posterior interval −3.8–6.2%) under model RHM-L-N, and 1.6% (interval −2.2–5.7%) under RHM-L-DP.

We performed several diagnostic assessments and sensitivity analyses to evaluate our model fit and to demonstrate the robustness of our results to model specification (see appendix F of the on-line supplementary materials for details). Though the within-county model (1) does not account for the potential for auto-correlation in the hospitalization time series, exploratory data analysis revealed little evidence of residual auto-correlation in our data. In particular, when we fitted model (1) separately for each county and inspected the auto-correlation function of the deviance residuals, we did not find a consistent pattern in the auto-correlation function. We further investigated whether there was spatial correlation across counties by plotting a variogram of the estimated county-specific θ_i, as well as whether there was residual spatial correlation in the county-specific estimates after accounting for long-term average NO_2 levels (appendix F in the on-line supplementary materials). We did not find evidence of spatial dependence across counties in the risk of cardio-vascular admissions associated with O_3 and $PM_{2.5}$ both above their national standards. To assess the sensitivity of our results to the specification of the expo-

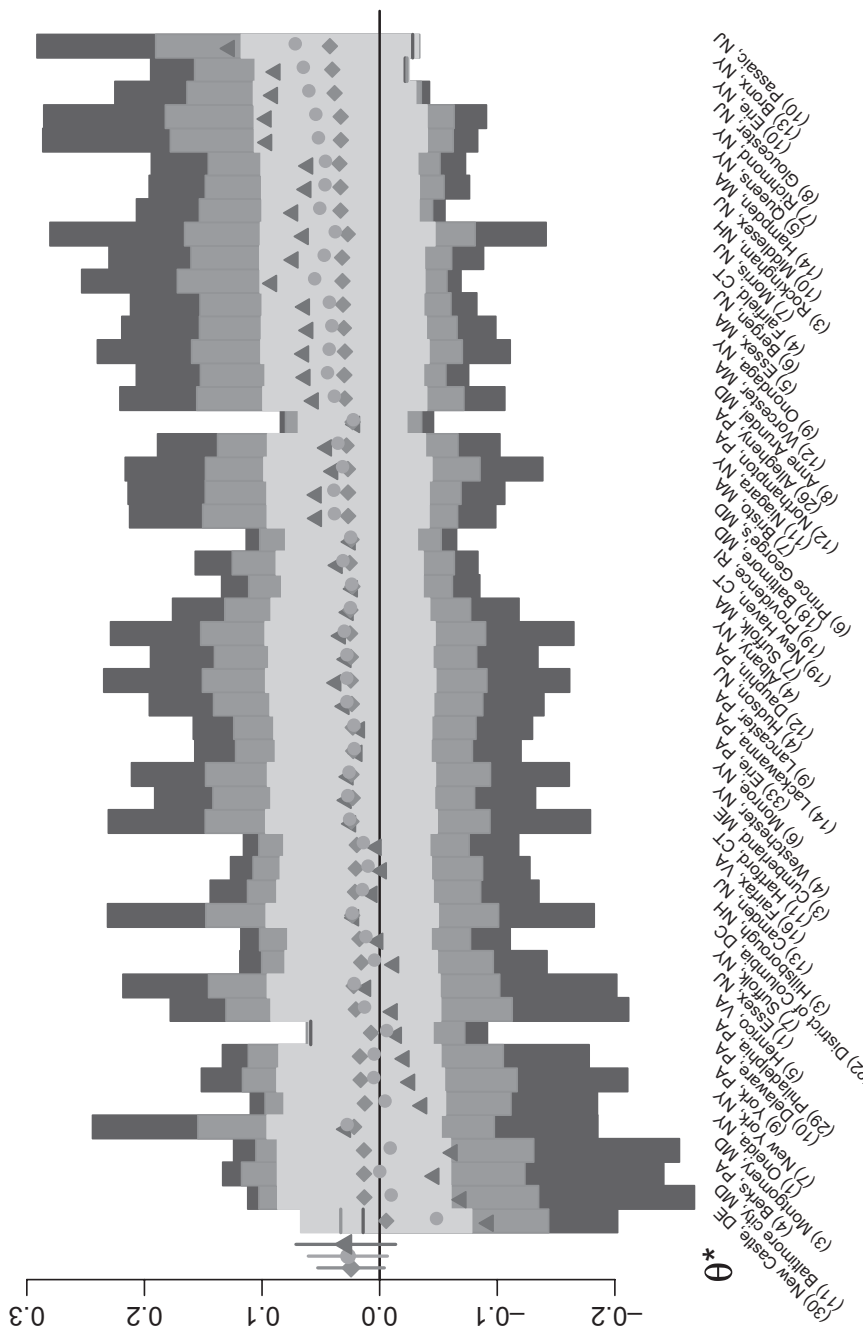

Fig. 3. Results of a multisite time series study of 51 north-eastern US counties, 1999–2005: county-specific (θ_i) and overall (θ^*) estimates, with 95% posterior intervals, of the log-relative-risk of cardio-vascular admissions on days when both O_3 and $PM_{2.5}$ exceed their national standard compared with days when both pollutants are below their standards, across three reduced HMs; normal approximation to the likelihood with normal random-effect distribution (model RHM-N-N; □, ◆) and integrated likelihood with normal (model RHM-L-N; □, ●) and flexible (model RHM-L-DP; ■, ▲) random-effect distributions); counties are ordered from left to right by increasing values of $\hat{\theta}_i/\mathrm{SE}_i$ where $\hat{\theta}_i$ is the maximum likelihood estimator and SE_i is the estimated standard error; the number of days with both O_3 and $PM_{2.5}$ greater than their national standards is listed beside each city

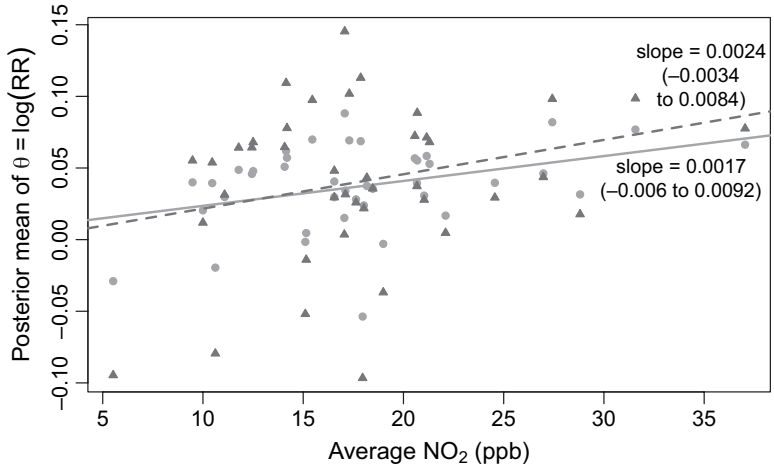

Fig. 4. For the 41 north-eastern US counties with NO_2 measurements, estimates of θ_i from the reduced HM incorporating long-term average NO_2 as a covariate in the second-stage model (●, model RHM-L-N; ▲, model RHM-L-DP): estimates of slopes α_1 (95% posterior intervals) are shown beside the corresponding trend line; the parameter of interest θ_i is the log-relative risk of cardio-vascular admissions on days when both O_3 level and $PM_{2.5}$ exceed their national standard compared with days when both are below their standards

sure–response surface, we refitted the reduced HM where the joint association of ozone and $PM_{2.5}$ with the health outcome in equation (1) was instead modelled as the product of cubic spline bases with just 2 DFs. We found that the resulting cluster-specific estimates θ_i were very similar and that the overall estimates θ^* were nearly identical.

7. Discussion

Although previous studies have estimated health effects of single pollutants, understanding how complex mixtures of pollutants affect health remains a challenging goal. Quantifying health risks resulting from exposure to a single pollutant is a useful analytical construct, but it is not representative of true exposure. It is therefore critical to develop models for estimating health effects of simultaneous exposure to multiple pollutants.

In this paper we developed methodology for estimating both county-specific and regional average risks of multipollutant exposure. This approach extends previous single-pollutant models by allowing for non-linear smooth functions of multiple pollutants and their interactions at the first stage and for effect modification at the second stage. Because flexible associations of several exposures are modelled concurrently, the inclusion of interactions of spline terms leads to a high dimensional vector of random effects. As a result, several challenges to the application of the usual full HM framework are introduced. To address these challenges, we have proposed the reduced HM as a general statistical approach for combining information across locations directly on the parameter of interest, in the context of many nuisance parameters. In this approach, information about the parameter of interest is summarized through a likelihood function (e.g. an integrated likelihood) in the first stage. At the second stage, a flexible random-effect distribution (e.g. a Dirichlet process normal mixture) is specified directly on the parameter of interest. We conducted simulation studies to compare performance of the reduced HM with the full HM, and we applied the reduced HM to a multisite time series study of 51 north-eastern US counties during the period 1999–2005.

In comparison with the reduced HM, on first inspection the full HM is the seemingly optimal approach, as it uses all of the available data in a single model to combine information across clusters. However, many practical difficulties may arise on implementation. First, for the full HM one must specify a random-effect distribution on the vector $\boldsymbol{\beta}_i$ parameterizing the within-cluster model. This may be difficult when the $\boldsymbol{\beta}_i$ are high dimensional or when they do not have meaningful interpretations (e.g. regression spline coefficients as in equation (1)). Additionally, for conducting Bayesian inference, prior distributions must be selected for the parameters of the random-effect distribution (e.g. the mean vector $\boldsymbol{\beta}^*$ and variance–covariance matrix Σ), which may also be complicated if these parameters do not have meaningful interpretations. If a reparameterization of $\boldsymbol{\beta}_i$ such that $\boldsymbol{\beta}_i = (\theta_i, \boldsymbol{\lambda}_i)$ for $\boldsymbol{\lambda}_i$ a $(q-1)$-dimensional nuisance parameter does not exist, then prior information about the quantity of interest $\theta_i = h(\boldsymbol{\beta}_i; \mathbf{x}_i)$ cannot be easily translated into prior information about the model parameters $\boldsymbol{\beta}_i$. Moreover, if one is interested in effect modification of cluster-specific covariates $\mathbf{z}_i$ at the second level, then a potentially high dimensional multivariate regression model for $\boldsymbol{\beta}_i | \boldsymbol{\beta}^*, \mathbf{z}_i$ must be specified. Finally, fitting the model (e.g. implementing the MCMC sampler) will become increasingly challenging and computationally intensive as the dimension of $\boldsymbol{\beta}_i$ (the number of random effects) increases.

For the reduced HM, however, rather than specify a high dimensional random-effect distribution on parameters that are not of primary scientific interest, one needs to specify only a random-effect distribution for a one-dimensional parameter that has a meaningful interpretation. Additionally, it is frequently much easier to incorporate prior information about the parameter of interest θ_i than about a large vector of nuisance parameters $\boldsymbol{\beta}_i$ that may be difficult to interpret (e.g. spline coefficients). Furthermore, reducing an HM on a high dimensional vector of parameters to an HM on a much lower dimensional space yields simpler implementation and greater computational efficiency, and makes model diagnostics and sensitivity analyses more wieldy.

Although the reduced HM overcomes many difficulties in the specification and implementation of the full HM, it also introduces new challenges. At the first stage, one must eliminate nuisance parameters to obtain the likelihood function $L_i(\theta_i)$. Although the literature on likelihood-based methods for eliminating nuisance parameters is vast (Pawitan, 2001; Edwards, 1992), in this paper we restricted our attention to those likelihoods that correspond to true probability distributions, including the integrated and conditional likelihood. In the case of large within-cluster sample sizes, the choice of which likelihood function to use should make little difference compared with the effect of the selection of the random-effect distribution. For smaller sample sizes, an integrated likelihood, though more computationally intensive than a normal approximation, allows greater flexibility for capturing the true form of the likelihood. Second, although the reduced HM avoids the need to specify a high dimensional random-effect distribution on the complex $\boldsymbol{\beta}_i$, use of the integrated likelihood for θ_i still necessitates specifying priors for $\boldsymbol{\beta}_i$ (Section 4.1). However, because we seek an objective likelihood function in the sense that it should summarize the information that is contained in the data about the parameter of interest such that the prior has as little influence as possible, any prior distribution for $\boldsymbol{\beta}_i$ that induces a vague prior for θ_i will suffice. For the applications that we have considered, assuming diffuse normal priors for each component of $\boldsymbol{\beta}_i$ leads to a prior for θ_i that is flat over a large range of reasonable values for θ_i, and we have found that this approach works well. Alternative approaches for approximating the likelihood function could also be considered, such as the data cloning method of Lele *et al.* (2007). Third, although one gains simplicity by eliminating nuisance parameters at the outset, it is possible that some information may be lost before combining information across clusters.

We conducted a series of simulation studies to evaluate the relative performance of the reduced HM compared with the full HM across a range of potential scenarios (Section 5). For the full HM, because one must specify random-effect distributions for a larger number of parameters, which may also be difficult to interpret, there is more potential for model misspecification than for the reduced HM where a random-effect distribution is placed on the lower dimensional parameter of interest. However, if the parameter of interest θ_i is correlated with nuisance parameters within a cluster, then information may be lost by reducing the parameter space to a single parameter and pooling the θ_i. We based the simulation study on an application for which a conditional likelihood for θ_i was available in closed form to focus on the effect on inference of misspecifying the random-effect distribution, rather than of misspecifying the likelihood function. In addition, though prior studies have considered the special case of reduced HMs where a conditional likelihood is available (Efron, 1996; Liao, 1999), the relative performance of this approach compared with the full HM had not been previously studied. When we refitted the reduced HM by using an integrated likelihood for a subset of the simulations ($I = 100$ and $n = 100$), we found that the performance for estimating the cluster-specific and overall parameters were either identical or just slightly worse than using the conditional likelihood. Across simulation scenarios, we found that the reduced HM generally achieved performance that was comparable with the full HM, and even had superior performance in some cases. We also performed a separate simulation study to evaluate the performance of our approach for estimating the integrated likelihood (Section 4.1) in a scenario based on our multipollutant application, finding that the estimated integrated likelihood closely matched the true integrated likelihood (appendix E of the on-line supplementary materials). Taken together, our findings from these simulation studies highlight the utility of the reduced HM both specifically to the multipollutant application and more generally to the context of two-level clustered data.

Development of reduced HMs was motivated by methodological needs for estimating health risks of joint exposure to multiple pollutants. We applied the reduced HM methodology to estimate the risk of emergency cardio-vascular admissions associated with simultaneous exposure to fine particulate matter and O_3. For the overall effect θ^*, we found marginal evidence of increased risk on days when both pollutants exceeded their national standards compared with when both were below their national standards. The reduced HM with normal random-effect distribution on the parameter of interest θ_i (model RHM-L-N) led to more shrinkage of the county-specific random effects than the reduced HM with flexible random-effect distribution (model RHM-L-DP). Further, RHM-L-N had narrower credible intervals for the county-specific parameters θ_i than RHM-L-DP. If the normal random-effect distribution is misspecified (e.g. if the analysis is missing an important county level effect modifier) then model RHM-L-N may understate statistical uncertainty in the θ_i. We illustrated how diagnostics on the reduced parameter space could be performed to assess modelling assumptions, by investigating spatial auto-correlation in the risk of simultaneous exposure to $PM_{2.5}$ and O_3. Though we did not find evidence of spatial auto-correlation in the θ_i in this application, it would be straightforward to model spatial dependence in the second stage of the reduced HM by specifying a spatial model for $\mathrm{cov}(\theta_i, \theta_j)$. We also demonstrated that the reduced HM can easily accommodate effect modifiers. Specifically, we examined the inclusion of long-term county level NO_2, which is a surrogate for traffic exposure. We found a larger relative risk of cardio-vascular admissions associated with levels of $PM_{2.5}$ and O_3 that are higher than their national standards in locations with high average NO_2 compared with locations with low average NO_2 levels, although the effect modification was not statistically significant. For our within-county model (1) and parameter of interest θ_i (which is defined in expression (2)) we considered only the association of current day's exposure to $PM_{2.5}$ and O_3 with hospitalization on the same day, though previous days'

exposure (e.g. at different lags from the present day) may also be predictive of health outcome. This choice of lag was motivated by previous single-pollutant studies, which have found that the strongest effects for $PM_{2.5}$ and O_3 occur at short (current or 1 day before) lags (Dominici *et al.*, 2006; Bell *et al.*, 2004). Furthermore, to demonstrate our methodology, we considered just a single example of a policy relevant parameter of interest. The US Environmental Protection Agency is considering introducing joint national standards to protect human health from the risks of exposure to complex mixtures better, and so studies providing a scientific basis for joint standards are needed (Dominici *et al.*, 2010). Depending on the scientific question, alternative parameters of interest may be specified and the same methodology applied. We could consider, for example, the gradient of the air pollution–hospitalization exposure–response surface at the national standards, or the relative risk of adverse health events when both $PM_{2.5}$ and O_3 exceed their national standard at different temporal lags compared with when just one of the pollutants exceeds its standard. In the future we shall apply this approach to conduct systematically a national investigation of the health effects that are associated with simultaneous exposure to multiple pollutants. Methods can be extended to an arbitrarily large number of pollution variables and locations, and to consider joint pollutant exposure at different lags as well as multiple parameters of interest that summarize different salient features of the multivariate exposure–response surface.

There are several extensions to the reduced HM methodology that we have proposed. First, we assumed a within-location model that had the same form across locations. However, this assumption could be relaxed. One could specify different within-cluster models for each cluster, as long as the interpretation of the parameter of interest remains constant across models. For example, for the within-cluster model (1) in the multipollutant application, the full HM would require a common spline basis (e.g. common knot locations) for the joint O_3 and $PM_{2.5}$ association across locations, whereas the reduced HM can allow for locally optimized spline bases. Thus the reduced HM approach can readily accommodate heterogeneity in the appropriate model to use across locations. In this paper we focused on two-level clustered data sets and a scalar parameter of interest. However, the reduced HM could be generalized to three- or higher level models, and to situations where the parameter of interest $\boldsymbol{\theta}_i = h(\boldsymbol{\beta}_i)$ is a multivariate parameter with $\dim(\boldsymbol{\theta}_i) < \dim(\boldsymbol{\beta}_i)$.

We have described the reduced HM within the context of estimating health risks of exposure to many pollutants. However, this hierarchical modelling strategy is broadly applicable to clustered data in which the parameter of interest is a known function of the vector of parameters $\boldsymbol{\beta}_i$ of the within-cluster model. The meta-analysis of stomach ulcer treatment that served as the basis for our simulation study is one example. Another example is the estimation of heat wave mortality risk in multisite time series studies (Bobb *et al.*, 2011). One can build a location-specific model that is similar to model (1) where the exposure–response function of interest is the temperature–mortality relation, adjusted for time varying covariates. One can then define a heat wave day indicator variable as a function of temperature on current and previous days. The parameter of interest θ_i, which is defined as the log-relative-risk of mortality on heat wave days compared with non-heat-wave days (see for example Peng *et al.* (2011)), can then be written as a known function of the temperature–mortality exposure–response function (parameterized by $\boldsymbol{\beta}_i$), and the reduced HM framework may be applied.

The reduced HM is especially useful in situations where $\boldsymbol{\beta}_i$ is high dimensional, where the components of $\boldsymbol{\beta}_i$ are not easily interpretable or where one wishes to incorporate prior information directly on the parameter of interest. For such applications, the reduced HM allows us to specify a random-effect distribution directly on the parameter of interest θ_i and to study effect modification by specifying an across-cluster regression model for θ_i. Further, the reduced

parameter space leads to simpler implementation, which facilitates the specification of flexible random-effect distributions that do not require strong assumptions on the random effects. For problems that are very high dimensional in the number of clusters, the number of observations within a cluster, and the number of parameters in the within-cluster model, it may not be computationally feasible to fit a full HM. In such cases, the reduced HM is a practical alternative.

8. Supplementary materials

The reader is referred to the on-line supplementary materials for technical appendices and additional simulation study results.

Acknowledgements

This work was supported by National Institute of Environmental Health Sciences grants T32ES012871, R01ES012054 and R01ES019560, and US Environmental Protection Agency grants RD-83479801 and RD-83241701.

References

Agresti, A., Caffo, B. and Ohman-Strickland, P. (2004) Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computnl Statist. Data Anal.*, **47**, 639–653.

Bell, M., McDermott, A., Zeger, S., Samet, J. and Dominici, F. (2004) Ozone and short-term mortality in 95 US urban communities, 1987-2000. *J. Am. Med. Ass.*, **292**, 2372–2378.

Berger, J., Liseo, B. and Wolpert, R. (1999) Integrated likelihood methods for eliminating nuisance parameters. *Statist. Sci.*, **14**, 1–28.

Bobb, J. F., Dominici, F. and Peng, R. D. (2011) A Bayesian model averaging approach for estimating the relative risk of mortality associated with heat waves in 105 U.S. cities. *Biometrics*, **67**, 1605–1616.

Chen, J., Zhang, D. and Davidian, M. (2002) A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, **3**, 347–360.

Dominici, F., Peng, R., Barr, C. and Bell, M. (2010) Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. *Epidemiology*, **21**, 187–194.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L. and Samet, J. M. (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Ass.*, **295**, 1127–1134.

Dominici, F., Samet, J. M. and Zeger, S. L. (2000) Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy (with discussion). *J. R. Statist. Soc.* A, **163**, 263–302.

Edwards, A. (1992) *Likelihood*. Baltimore: Johns Hopkins University Press.

Efron, B. (1996) Empirical Bayes methods for combining likelihoods. *J. Am. Statist. Ass.*, **91**, 538–565.

Environmental Protection Agency (2011) The benefits and costs of the Clean Air Act from 1990 to 2020. *Technical Report*. US Environmental Protection Agency Office of Air and Radiation, Washington DC.

Ferguson, T. (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

Gallant, A. R. and Nychka, D. W. (1987) Semi-nonparametric maximum likelihood estimation. *Econometrica*, **55**, 363–390.

Gilks, W., Richardson, S. and Spiegelhalter, D. (eds) (1995) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Heagerty, P. and Kurland, B. (2001) Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, **88**, 973–985.

Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Tertre, A. L., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F. and Boumghar, A. (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the aphea2 project. *Epidemiology*, **12**, 521–531.

Komárek, A. and Lesaffre, E. (2008) Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Computnl Statist. Data Anal.*, **52**, 3441–3458.

Laird, N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *J. Am. Statist. Ass.*, **73**, 805–811.

Lele, S. R., Dennis, B. and Lutscher, F. (2007) Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. *Ecol. Lett.*, **10**, 551–563.

Liao, J. (1999) A hierarchical Bayesian model for combining multiple 2 × 2 tables using conditional likelihoods. *Biometrics*, **55**, 268–272.
Litière, S., Alonso, A. and Molenberghs, G. (2008) The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statist. Med.*, **27**, 3125–3144.
Litière, S., Alonso, A. and Molenberghs, G. (2010) Rejoinder to "A note on type II error under random effects misspecification in generalized linear mixed models". *Biometrics*, **66**, 656–660.
Magder, L. S. and Zeger, S. L. (1996) A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J. Am. Statist. Ass.*, **91**, 1141–1151.
McCulloch, C. E. and Neuhaus, J. M. (2011) Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics*, **67**, 270–279.
Metropolis, N., Rosenbluth, A., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.
National Research Council (2004) Research priorities for airborne particulate matter: IV, Continuing research progress. *Technical Report*. National Research Council of the National Academies, Washington DC.
Neal, R. (2000) Markov chain sampling methods for Dirichlet process mixture models. *J. Computnl Graph. Statist.*, **9**, 249–265.
Neuhaus, J., Hauck, W. and Kalbfleisch, J. (1992) The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, **79**, 755–762.
Pawitan, Y. (2001) *In All Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford: Oxford University Press.
Peng, R. D., Bell, M. L., Geyh, A. S., McDermott, A., Zeger, S. L., Samet, J. M. and Dominici, F. (2009) Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environ. Hlth Perspect.*, **117**, 957–963.
Peng, R. D., Bobb, J. F., Tebaldi, C., McDaniel, L., Bell, M. L. and Dominici, F. (2011) Toward a quantitative estimate of future heat wave mortality under global climate change. *Environ. Hlth Perspect.*, **119**, 701–706.
Peng, R., Chang, H., Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M. and Dominici, F. (2008) Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. *J. Am. Med. Ass.*, **299**, 2172–2179.
Peng, R. D., Dominici, F. and Louis, T. A. (2006) Model choice in time series studies of air pollution and mortality (with comments). *J. R. Statist. Soc.* A, **169**, 179–203.
Schwartz, J. and Coull, B. A. (2003) Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics*, **4**, 539–553.
Smith, T., Spiegelhalter, D. and Thomas, A. (1995) Bayesian approaches to random effects meta-analysis: a comparative study. *Statist. Med.*, **14**, 2685–2699.
Verbeke, G. and Lesaffre, E. (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computnl Statist. Data Anal.*, **23**, 541–556.
Warn, D. E., Thompson, S. G. and Spiegelhalter, D. J. (2002) Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statist. Med.*, **21**, 1601–1623.
Welty, L. J. and Zeger, S. L. (2005) Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season?: a sensitivity analysis using flexible distributed lag models. *Am. J. Epidem.*, **162**, 80–88.
Zeka, A. and Schwartz, J. (2004) Estimating the independent effects of multiple pollutants in the presence of measurement error: an application of a measurement-error-resistant technique. *Environ. Hlth Perspect.*, **112**, 1686–1690.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary Materials for "Reduced hierarchical models with application to estimating health effects of simultaneous exposure to multiple pollutants"'.